A Deep Learning Approach for Parkinson's Disease Diagnosis from EEG Signals

Article *in* Neural Computing and Applications · August 2020

DOI: 10.1007/s00521-018-3689-5

A Deep Learning Approach for Parkinson's Disease Diagnosis from EEG Signals

Shu Lih Oh[a], Yuki Hagiwara[a], U Raghavendra[b], Rajamanickam Yuvaraj[c], N Arunkumar[d], Murugappan M[e], U Rajendra Acharya[a,f,g,*]

[a]Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore.
[b]Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, India.
[c]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.
[d]Department of Electronics and Instrumentation, SASTRA University, Thanjavur
[e]Kuwait College of Science and Technology, Kuwait.
[f]Department of Biomedical Engineering, School of Science and Technology, Singapore University of Social Sciences, Singapore.
[g]School of Medicine, Faculty of Health and Medical Sciences, Taylor's University, 47500 Subang Jaya, Malaysia.
[*]Postal Address: Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore 599489.
Telephone: +65 6460 6135; Email Address: aru@np.edu.sg

Abstract

An automated detection system for Parkinson's disease (PD) employing the convolutional neural network (CNN) is proposed in this study. PD is characterized by the gradual degradation of motor function in the brain. Since it is related to the brain abnormality, electroencephalogram (EEG) signals are usually considered for the early diagnosis. In this work, we have used the EEG signals of twenty PD and *twenty* normal subjects in this study. A *thirteen*-layer CNN architecture which can overcome the need for the conventional feature representation stages is implemented. The developed model has achieved a promising performance of 88.25% accuracy, 84.71% sensitivity, and 91.77% specificity. The developed classification model is ready to be used on large population before installation of clinical usage.

Keywords – Computer-aided detection system; convolutional neural network; deep learning; Parkinson's disease.

1. Introduction

The human brain, at the time of birth, contains the maximum number of nerve cells also called neurons [1]. These nerve cells cannot get fixed on their own as the other cells of our body. With age, the neurons die out and hence become irreplaceable [2]. PD typically arises with the death of neurons [3]. The neurons produce a chemical substance known as dopamine and the main function of this is to control the movement of the body. Hence, as neurons die, the amount of dopamine produced in the brain decreases. As a result, this neurological condition starts to take place very slowly and influences various communication modes in the brain [4]. It has been observed that people around the age of 50 or older have been diagnosed with PD. The primary symptoms of this disease are unstable posture, stiffness in the muscle, slow movements, tremor, loss of balance and damaged fine motor skill [4].

This disease has taken over almost 10 million people based on the statistics provided by the World Health Organization [5]. There have been difficulties in diagnosing the disease when no apparent motor or non-motor symptoms were observed. Therefore, computer-aided diagnosis (CAD) system may be able to help in the early detection of any abnormalities [6, 7]. The CAD system is an automated detection system that can objectively diagnose PD using electroencephalogram (EEG) signals. The functions of the cortical and subcortical parts of the brain are easily identified with the help of the EEG. The neurological diseases like epilepsy, schizophrenia, Alzheimer's can also be determined using the EEG signals [8-11]. Therefore, we have used EEG signals to develop the CAD system for the detection of PD in this study.

According to previous studies, EEG signals are complex and non-linear in nature and hence, many linear feature extraction approaches are unable to accurately characterize these signals [6]. When the EEG signal displays complexity, aggravation of the PD is observed. This is due to the non-linear components present in the EEG signals [12-15]. Hence, it can be noted that the employment of nonlinear features extraction techniques would be useful in the differentiation of normal and PD EEG signals.

However, a sub-division of machine learning called deep learning has been successfully implemented in diverse areas of pattern recognition and the processing of natural language in recent years [16]. The convolutional neural network (CNN) is one of the most popular forms of deep learning that researchers adopted [17-23]. It allows the learning of higher level features without human intervention through training of the data, unlike most traditional machine learning algorithms. To the best of our knowledge, this is the first paper to implement the deep CNN for the CAD system for PD. We have implemented a novel *thirteen*-layer deep CNN to characterize the two classes (PD and normal). Figure 1 illustrates the architecture of the proposed network. The details of the network and each layer are presented in the subsequent sections.

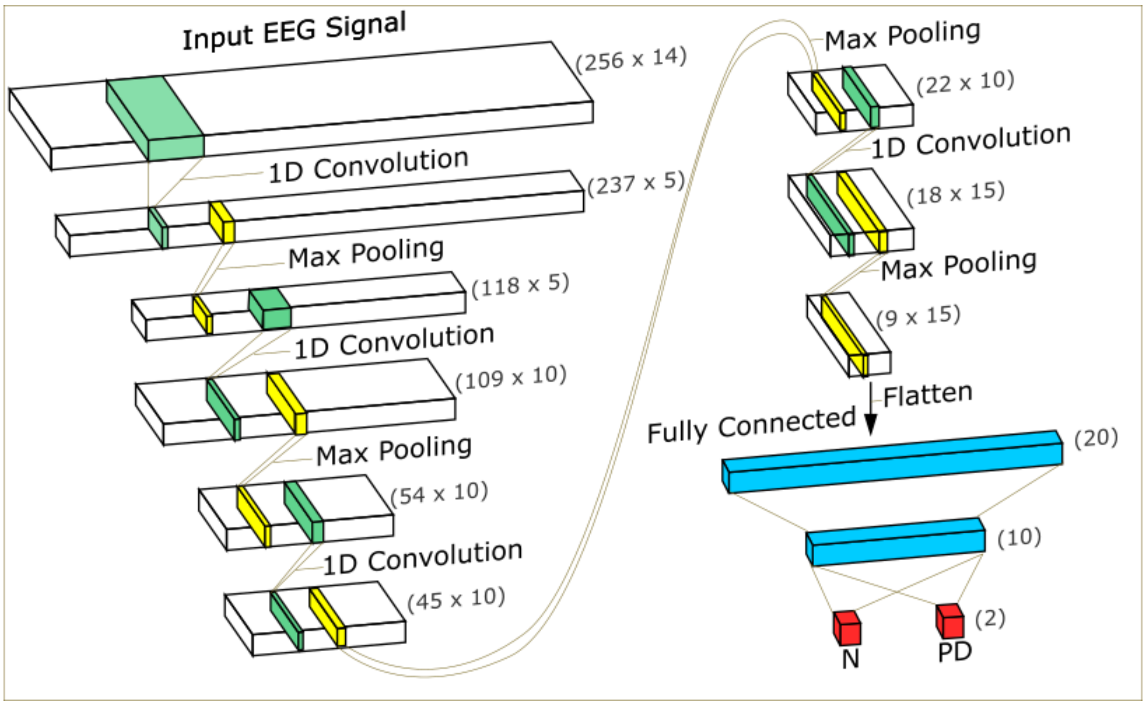


Figure 1 The proposed CNN architecture.

2. Deep Learning

It is a type of machine learning which effectively combines both feature extraction and classification processes [17-20]. Features extracted from the input data is used to build a robust CNN model, and afterward to test the diagnosis performance of the developed model during the testing phase. The CNN model has been successfully developed in the recent papers of Acharya et al. [21, 22] to automatically detect seizure and depression respectively using EEG signals.

2.1 Convolution Neural Network

The basic layers of the CNN include the convolution, max-pooling and fully-connected (dense) layer [23, 24]. Typically, the network tends to learn better as the network gets deeper [24].

However, this may affect the computational time. Hence, we have carefully designed the network architecture which requires shorter computational time. The highest classification performance is obtained with parameters which are finely tuned during the training phase.

The convolutional layer convolves with the input signal using a kernel (window) [25, 26]. A feature map for the next layer is generated after the convolution. After which, the batch normalization layer is applied to normalize the input training data to flow between the intermediate layers. The purpose is to enable faster learning and boosting. Then, the rectified linear unit is applied to threshold the input data and reduced the redundancies in the data. To reduce the size of the feature map, the max-pooling layer is used. Finally, every neuron of the max-pooling layer is connected to every neuron in the fully-connected layer where the output predicts the outcome (normal or PD) of the input signal [27, 28].

2.2 Proposed CNN Architecture

The overview of the proposed architecture is given in Figure 1. The validation of the model is carried out in two-steps. First, training the data and then testing the model. During training, stratified ten-fold cross-validation is introduced, where the full data is split into 10 uniform portions. 9 out of 10 parts are used to train the model whereas the rest are kept for testing. This procedure is iterated ten times so that all the ten parts will be involved in both the training and testing phases. Second, in order to assess the progression of the training at the end of each epoch, 20% of the cross-validation training data is allocated for validating the model. Adam optimization [29] with a learning rate of 0.0001 is used and we have used a few activation functions such as Relu for all layers and softmax for the last layer. The dropout is set to 0.5 for dropout layer. Table 1 shows the internal details of all layers. All the parameters are tuned accordingly to the training set provided that gives the optimum training accuracy. The kernel size and number of filters are obtained through the brute force technique.

Table 1 Details of parameters belonging to different layers of the developed CNN model.

Layers	Layer Name	Kernel Size	No. Of Filters	Stride	Output Shape	No. Of Trainable Parameters	Regularization
0	Input	-	-	-	256 x 14	0	-
1	1D convolution	20 x 1	5	1	237 x 5	1400	-
2	Max-pooling	2 x 1	5	2	118 x 5	0	-
3	1D convolution	10 x 1	10	1	109 x 10	500	-
4	Max-pooling	2 x 1	10	2	54 x 10	0	-
5	1D convolution	10 x 1	10	1	45 x 10	1000	-
6	Max-pooling	2 x 1	10	2	22 x 10	0	-
7	1D convolution	5 x 1	15	1	18 x 15	750	-
8	Max-pooling	2 x 1	15	2	9 x 15	0	-
10	Dense	-	-	-	20	2720	Dropout (0.5)
12	Dense	-	-	-	10	210	Dropout (0.5)
13	Dense	-	-	-	2	22	-
					Total	6602	

3. Experimental Results

3.1 Normal and PD Subjects

The EEG signals of 20 PD patients (10 women and 10 men) with an age range between 45 to 65 years old were collected upon the approval from the Hospital University Kebangsaan Malaysia Ethics Committee. These patients have an average period of PD of 5.75 ± 3.52 years (ranging from 1-12 years). The Hoehn and Yahr stages [30] are as follows:

Stage i) had 2,

Stage ii) had 11 and

Stage iii) had 7 PD patients

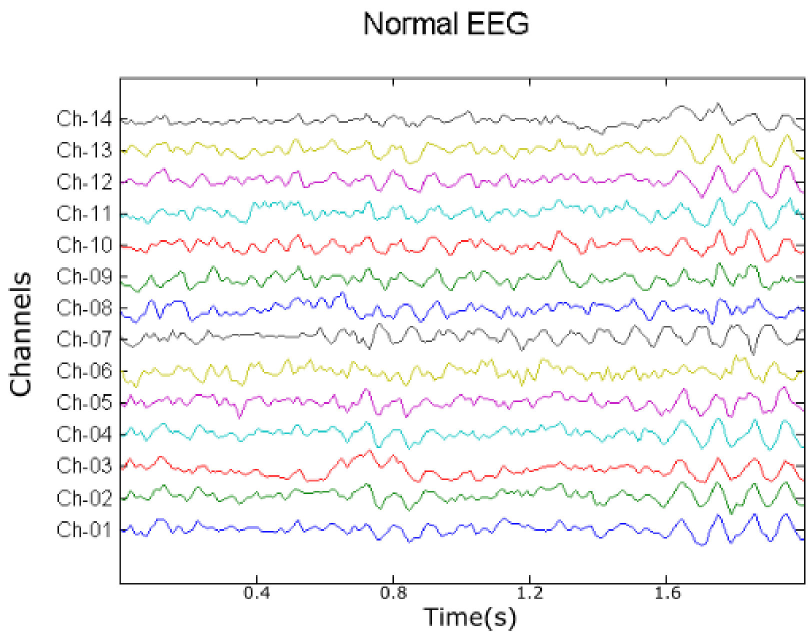
The mini-mental status examination (MMSE) results were observed to be inside the range of the normal limits (26.90 ± 1.51 [range 25 - 30]). The exclusion specification involves the existence of additional neurological conditions like epilepsy or psychiatric disorders such as depression and other acute mental disorder. Levodopa (L-dopa) drugs were consumed by PD patients to reduce the non-uniformity in the medication.

A total of 20 normal subjects in the same age group (9 men and 11 women with a mean age of 58.10 ± 2.95) with no past record or indications of neurological or mental disorder were enlisted. The MMSE results for the healthy participants were 27.15 ± 1.63 years. The normal and PD subjects are right-handed and were verified by the Edinburgh Handedness Inventory. Also, these subjects declared to have perfect hearing conditions. The participant's consent was sought for the study by explaining to them the probability of the risks involved.

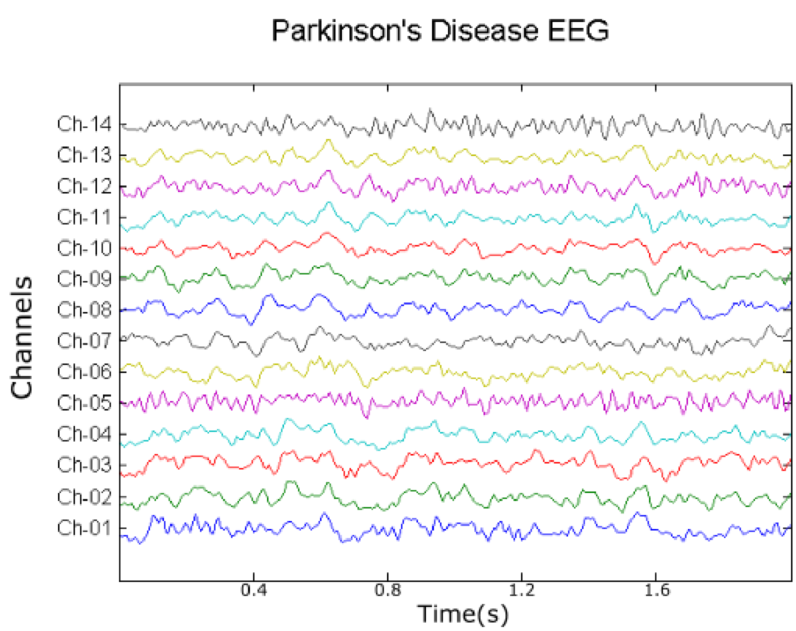
3.2 EEG Recordings and Pre-Processing

The recordings lasted 5 minutes in resting state (to attain a state of relaxed wakefulness) at 128 Hz sampling rate. An emotive EPOC neuroheadset of 14 channels was used. The participants were asked to sit comfortably in a quiet room and were informed before the recording to refrain from body movements (e.g., blinking of eyes) during the recording session. After the recording, the signals were segmented into 2-seconds window length.

A threshold technique was used to discard the signal amplitudes exceeding ± 100 μV to remove the eye blinking artifacts. Then, a 6^{th} order bandpass Butterworth filter with forward reverse filtering technique was employed to filter the frequency range of 1 – 49 Hz. Finally, 1588 artifact-free epochs were processed for further analysis. Figure 2 shows a sample of normal and PD EEG recordings.



(a)



(b)

Figure 2 A sample of (a) normal and (b) PD EEG signal.

3.3 Results

All the EEG signals were subjected to the proposed CNN model. The CNN network was designed in Python language using Keras and was executed on a computer with a system configuration of two Intel Xeon 2.40 GHz (E5620) processors with a 24 GB random access memory.

The evaluation parameters namely the accuracy, sensitivity, and specificity were used. The best diagnostic performance is achieved with the learning rate of 0.0001. The proposed CNN model yielded an accuracy of 88.25%, sensitivity, and specificity of 84.71% and 91.77% respectively. Figure 3 and 4 shows the performance of the model with and without dropout layer respectively. It can be noted that without the dropout layer, there is a possibility of overfitting of data. In Figure 3, the accuracy of the training set does not differ much from the accuracy of the validation set whereas, in Figure 4, the accuracy of the validation set performs a lot worse as compared to the training data.

Figure 5 shows the confusion matrix of our results. It can be observed that 11.34% of normal subjects are misclassified as PD and 11.51% of the PD EEG signals are wrongly categorized into the normal class.

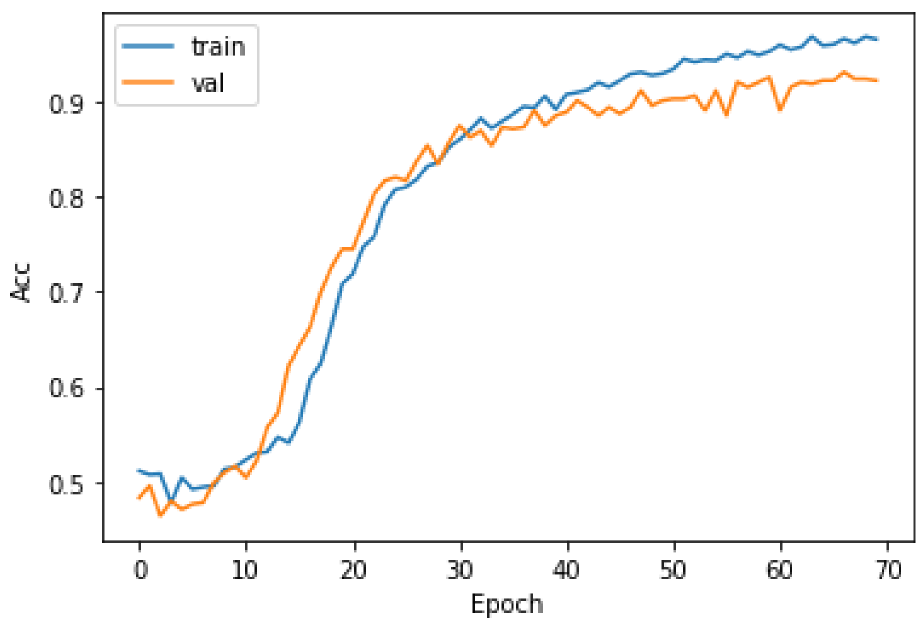


Figure 3 Accuracy versus different epoch plot.

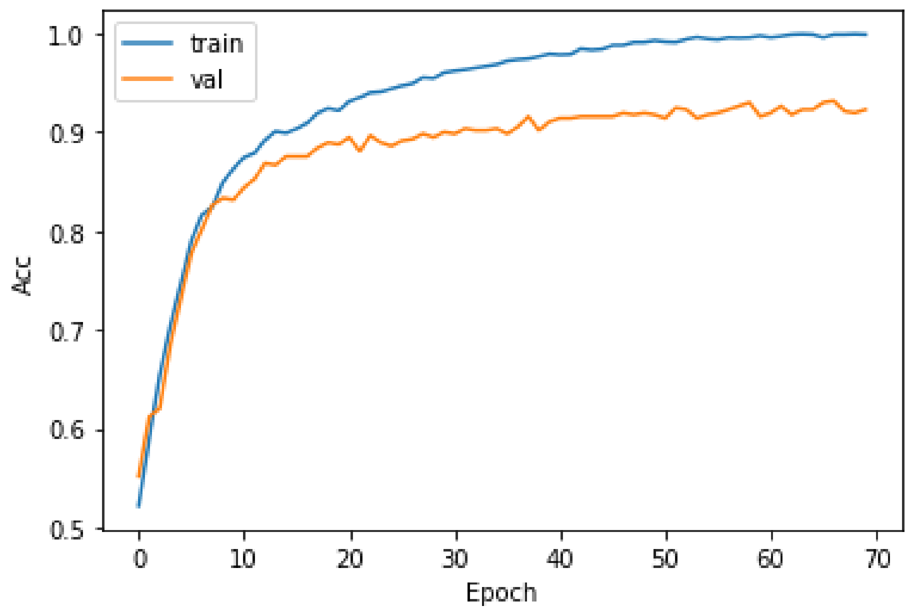


Figure 4 Accuracy versus different epoch without dropout layer plot.

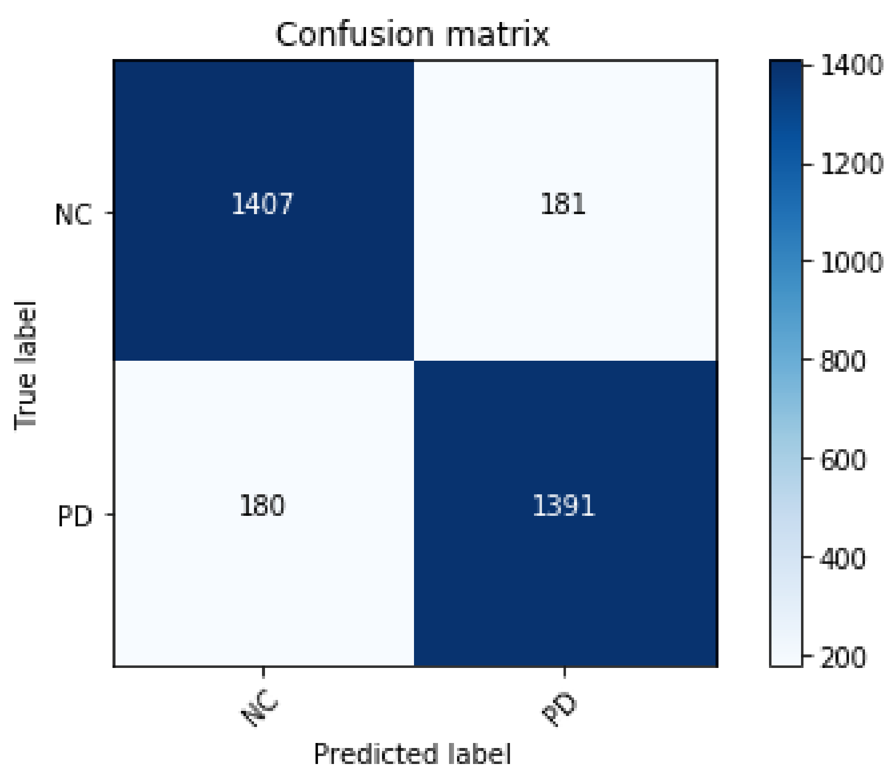


Figure 5 Confusion matrix of the proposed method.

4. Discussion

In the past several years, several non-invasive techniques have been proposed to identify PD using voice [31-33] and gait [34] signals. Different automated techniques were explored to develop the best automated model to differentiate normal and PD subjects. Chen et al. [31] suggested that feature reduction method could eliminate the unwanted information from the PD voice signals. Moreover, they have reported an average diagnostic accuracy of 96.07% with the principal component analysis feature reduction technique and the fuzzy k-nearest neighbor (FKNN) classifier. Later, the accuracy was improved by Zuo et al. [32] using swarm intelligence algorithm supplemented with FKNN classifier in classifying normal and PD voice signals. In addition, Ma et al. [33] proposed a combination of clustering algorithm with a kernel-based extreme learning machine classifier to characterize PD voice signal from normal voice signal and reported an average accuracy of 99.49%. On the other hand, Daliri [34] achieved a classification accuracy of 91.20% by using feature discriminant ratio based on Fourier Transform in the differentiation of normal and PD gait signals.

However, the number of studies using EEG signals to diagnose PD are limited. Based on the studies conducted (see Table 2), it can be noted that various machine learning techniques were adopted to discriminate the EEG signals in normal and PD subjects. Han et al. [6] conducted an experiment to investigate the characteristics of normal and PD EEG signals. They reported that the entropy values of PD EEG signals were significantly higher than normal EEG signals. This showed that the PD EEG signals are more complex. Yuvaraj et al. [7] employed higher-order statistics (HOS) feature extraction technique to differentiate the two classes of signals. It was reported that the HOS can explicitly represent the concealed nonlinear traits of the PD EEG signals for classification.

Nevertheless, in this study, we have proposed a deep CNN architecture to detect PD. The novelty of this method is the formulation of a *thirteen*-layer network to differentiate between PD and normal subjects using EEG signals. Furthermore, no hand-picked features are required in this study. This significantly minimizes the process of experimenting and selecting the best set of features for classification.

In addition, to further improve the efficiency of the CAD system, we have proposed a web-based diagnosis technique that could be initiated in the future. Figure 6 describes the workflow of the web-based CAD system. This system uses the internet to diagnose the PD patients. The EEG signals collected from patients are stored in the local storage server in the clinic and sent through the cloud where our developed CNN model is placed. The diagnosis is sent back from the cloud to the clinic. Furthermore, the advantage of this web-based application is that the diagnosis can be sent directly to the patient via text message. Hence, with the installation of this system, clinicians in the clinics will be able to significantly reduce their workload.

The main advantages of this proposed technique are:

I. A *thirteen*-layer CNN model is designed to automatically identify PD using EEG signals.
II. Extraction, selection, and classification of features are not required in the proposed CNN model.
III. The model is validated with a stratified ten-fold cross-validation technique.
IV. This the first work to implement the deep learning technique for the detection of PD using EEG signals.
V. Obtained good performance even with less number of normal and PD subjects. Hence, the developed is robust.

The main disadvantages of this proposed technique are:

I. Used limited number (20 normal and 20 PD) subjects to develop the CNN model.
II. The CNN structure is computationally expensive as compared to the conventional machine learning techniques.

Table 2 The summary of CAD system developed using EEG signals to diagnose PD.

Authors	Year	Techniques	Performances (%)		
			Accuracy	Sensitivity	Specificity
Han et al. [6]	2013	• Wavelet packet entropy • AR Burg	-	-	-
Yuvaraj et al. [7]	2016	• HOS and bi-spectrum features • Support vector machine classifier • Parkinson's disease diagnosis index	99.62	100.00	99.25
Present work	2018	• *Thirteen*-layer CNN	88.25	84.71	91.77

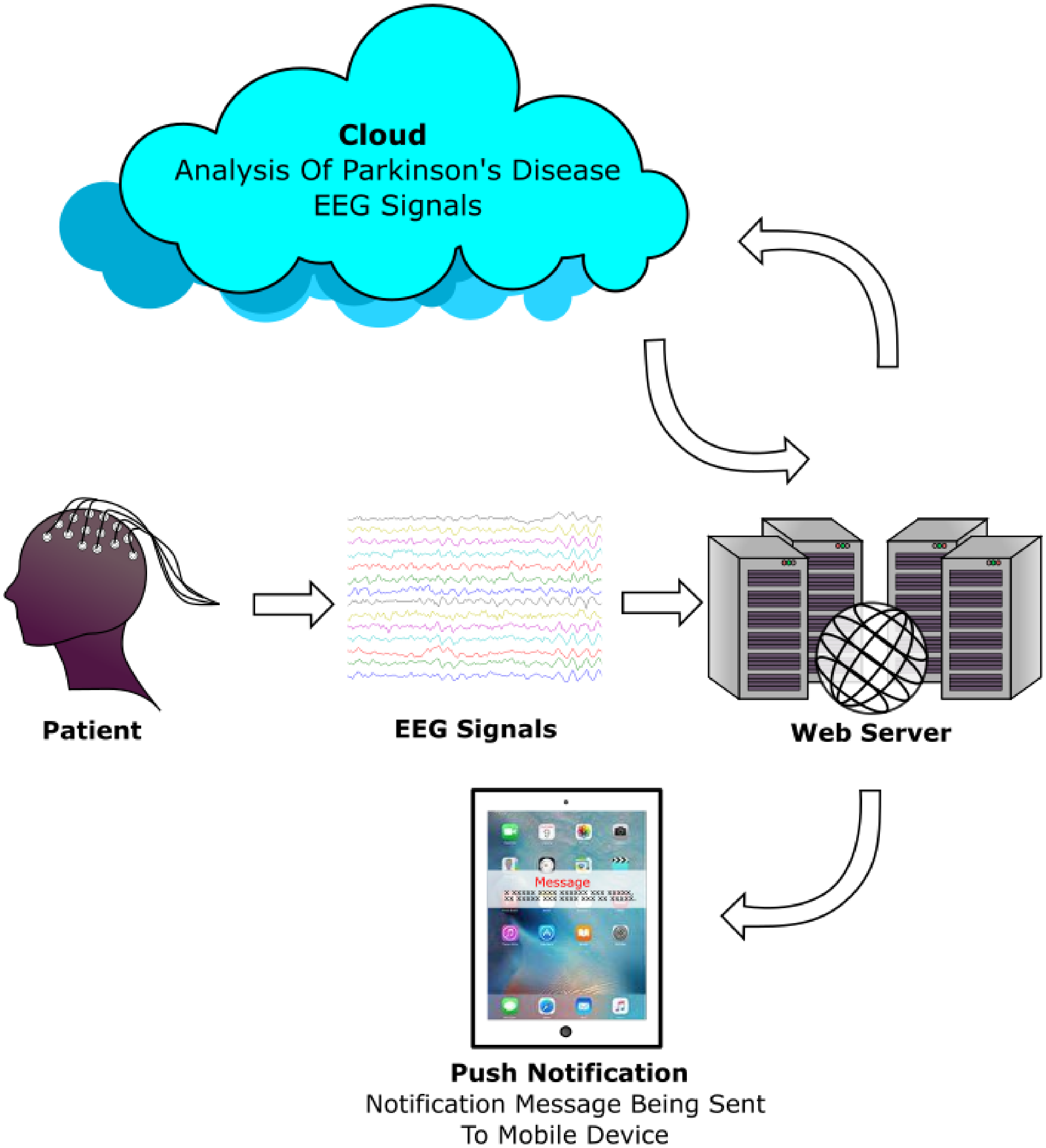


Figure 6 Web-based CAD system to diagnose PD.

In future, authors intend to use the developed model to use with huge database from belonging to different race and age groups. Also, such techniques can be used to detect other brain abnormalities like autism, Alzheimer's disease, depression and sleep disorders.

5. Conclusion

An automated *thirteen*-layer CNN model to diagnose PD using EEG signals is proposed. Furthermore, this is the first study which implemented the deep learning concept to diagnose the PD using EEG signals. We have obtained an accuracy of 88.25%, sensitivity of 84.71%, and specificity of 91.77% despite the limited number of subjects. Based on the positive performances achieved, the presented model may be able to serve as a trusted and long-term tool to assist clinicians in PD diagnoses. In future, authors propose to test the develop model with huge number of subjects and also aim to detect the early stage of PD.

6. Conflict of Interest

The authors declared no conflict of interest in this work.

7. References

1. Stiles J, Jernigan TL (2010) The basics of brain development. Neuropsychology Review 20(4):327-348
2. Silver J, Schwab ME, Popovich PG (2015) Central nervous system regenerative failure: role of oligodendrocytes, astrocytes, and microglia. Cold Spring Harbor Perspective in Biology 7(3):a020602
3. Surmeier DJ, Guzman JN, Sanchez-Padilla J, Goldberg JA (2010) What causes the death of dopaminergic neurons in Parkinson's disease? Progress in Brain Research 183:59-77
4. Heyn SN, Davis CP, Stoppler MC (2018) Parkinson's disease symptoms, signs, causes, stages, and treatment. Retrieved from Medicine Net, https://www.medicinenet.com/parkinsons_disease/article.htm#parkinsons_definition_and_disease_facts
5. World Health Organization, Neurological Disorders public health challenges (2006). Retrieved from http://www.who.int/mental_health/neurology/neurological_disorders_report_web.pdf

6. Han CX, Wang J, Yi GS, Che YQ (2013) Investigation of EEG abnormalities in the early stage of Parkinson's disease. Cognitive Neurodynamics 7:351-359
7. Yuvaraj R, Acharya UR, Hagiwara Y (2016) A novel Parkinson's diagnosis index using higher-order spectra features in EEG signals. Neural Computing and Applications, Springer London
8. Lima CAM, Coelho ALV, Chagas S (2009) Automatic EEG signal classification for epilepsy diagnosis with Relevance Vector Machines. Expert Systems with Applications 36(6):10054-10059
9. Leuchter AF, Cook IA, Gilmer WS, Marangell LB, Burgoyne KS, Howland RH, Trivedi MH, Zisook S, Jain R, Fava M, Iosifescu D, Greenwald S (2009) Effectiveness of a quantitative electroencephalographic biomarker for predicting differential response or remission with escitalopram and bupropion in major depressive disorder. Psychiatry Research 169(2):132-138
10. Gandal MJ, Edgar JC, Klook K, Siegel SJ (2012) Gamma synchrony: towards a translational biomarker for the treatment – resistant symptoms of schizophrenia. Neuropharmacology 62(3):1504-1518
11. Hampal H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, Herholz K, Bokde AL, Jessen F, Hoessler YC, Sanhai WR, Zetterberg H, Woodcock J, Blennow K (2010) Biomarker for Alzheimer's disease: academic, industry and regulatory perspectives. Nature Reviews Drug Discovery 9(7):560-574
12. Chua KC, Chandran V, Acharya UR, Lim CM (2009) Analysis epileptic EEG signals using higher order spectra. Journal of Medical Engineering and Technology 33(1):42-50
13. Acharya UR, Chua EC, Chua KC, Min LC, Tamura T (2010) Analysis and automatic identification of sleep stages using higher order spectra. International Journal of Neutral Systems 20(6):509-521
14. Martis RJ, Acharya UR, Mandana KM, Ray AK, Chakraborty C (2013) Cardiac decision making using higher order spectra. Biomedical Signal Processing and Control 8:193-203
15. Yuvraj R, Murugappan M, Norlinah MI, Sundaraj K, Omar MI, Khairiyah M, Palaniappan R (2014) Optimal set of EEG features for emotional state classification and trajectory visualization in Parkinson's disease. International Journal of Psychophysiology 94(3):482-495
16. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 1097-1105

17. Acharya UR, Fujita H, SL Oh, Hagiwara Y, Tan JH, Muhammad A (2017) Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Information Sciences 405:81-90
18. Acharya UR, Fujita H, Oh SL, Muhammad A, Tan JH, Chua KC (2017) Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. Knowledge-Based Systems 132: 62-71
19. Tan JH, Acharya UR, Bhandary SV, Chua KC, Sivaprasad S (2017) Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. Journal of Computational Science 20:70-79
20. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Muhammad A, Tan RS (2018) Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Applied Intelligence, Springer US, 1-12
21. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H (2017) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in Biology and Medicine DOI: https://doi.org/10.1016/j.compbiomed.2017.09.017
22. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H, DP Subha (2018) Automated EEG-based screening of depression using deep convolutional neural network. Computer Methods and Programs in Biomedicine 161:103-113
23. Acharya UR, Fujita H, Oh SL, Raghavendra U, Tan JH, Muhammad A, Gerytch A, Hagiwara Y (2018) Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. Future Generation Computer Systems 79(3):952-959
24. Faust O, Hagiwara Y, Tan JH, Oh SL, Acharya UR (2018) Deep learning for healthcare applications based on physiological signals: A review. Computer Methods and Programs in Biomedicine 161:1-13
25. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nature Neuroscience 2:1019–1025
26. Cireşan D, Meier U, Masci J (2011) A committee of neural networks for traffic sign classification. In 396 Proceedings of the 2011 International Joint Conference on Neural Networks, IEEE: California 397:1918-1921

27. Scherer D, Müller A, and Behnke S (2010) Evaluation of pooling operations in convolutional architectures for object recognition. In International conference on artificial neural networks Springer 82–91.
28. Serre T, Wolf L, and Poggio T (2005) Object recognition with features inspired by visual cortex. In Computer Vision and Pattern Recognition Conference 994–1000
29. Kingman DP, Ba J (2015) Adam: A method for stochastic optimization. 3rd International Conference for Learning Representations, San Diego
30. Yahr M, Hoehn M (1967) Parkinsonism: onset, progression and mortality. Neurology 17(5):427-442
31. Chen HL, Huang CC, Yu XG, Xuc X, Sund X, Wang G, Wang SJ (2013), An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Systems with Applications 40(1):263-271
32. Zuo WL, Wang ZY, Liu T, Chen HL (2013) Effective detection of Parkinson's disease using an adaptive fuzzy K-nearest neighbor approach. Biomedical Signal Processing and Control 8(4):364-373
33. Ma C, Ouyang J, Chen HL, Zhao XH (2014) An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. Computational and Mathematical Methods in Medicine 2014:1-14
34. Daliri MR (2013) Chi square distance kernel of the gaits for the diagnosis of Parkinson's disease. Biomedical Signal Processing and Control 8(1):66-70